

## A Goal-oriented Navigation Model based on Multi-scale Place Cells Map

Jia Du<sup>1,2,\*</sup>, Dewei Wu<sup>1</sup>, Weilong Li<sup>1</sup> and Yang Zhou<sup>1</sup>

<sup>1</sup> Information and Navigation College, Air Force Engineering University, Xi'an 710077, China

<sup>2</sup> Xi'an Communications Institute, Xi'an 710106, China  
E-mail [dujia0227@126.com](mailto:dujia0227@126.com)

### Abstract

*To achieve spatial cognition and autonomous navigation for the robot, learning from the biological mechanism for navigation, a goal-oriented navigation model based on multi-scale place cells map is proposed. First, a place cell map with different scale is constructed for coding the space environment. Second, the Gaussian model is used to simulate the firing rate of a place cell which will be the input of the Q-learning process. Third, the annealing strategy is explored to choose a reasonable action and update the Q table. After training and learning, the robot could succeed to finish the goal-oriented navigation task. Simulation results show that, the proposed method is feasible to complete the goal-oriented navigation task. Compared with the spatial cognitive model of single scale place cells, the method not only meets the multi-scale spatial representation nature of place cells in hippocampus, but also has a faster learning speed. The achievement of this paper has a reference for the study of brain-inspired navigation.*

**Keywords:** brain-based navigation; place cell; reinforcement learning; multi-scale representation

### 1. Introduction

As early as 1948, Tolman proposed the concept of cognitive map [1], and he considered that the rat formed a map in the brain characterizing the space environment and used this cognitive map to solve the problem of navigation. After three decades, O'Keefe found that, in the way of electrophysiology, the cognitive map like Tolman located in the hippocampus of the rat's brain [2]. Further, He found a special neural cell of which the firing intensity has a strong correlation with the position of the rat and is then called place cell. With the discoveries of the head directed cells[3][4], grid cells[5][6], the boundary cells[7], the speed of cell[8] in the brain, the mechanism of the biological brain navigation has been further developed. The combined effects of these cells form a sophisticated navigation system of the brain, which is like a GPS system that can provide the space navigation information in real time and finish the related tasks of navigation.

For the development of biological navigation mechanism in the field of brain science, the scholars in the navigation area are more interested in how these important discoveries are used to solve the actual navigation problems. Currently, there are also many researchers that have done some related work [9][13]. Reference [14] introduces the neural mechanisms of the biological navigation and discusses how the navigation cells encode the spatial distance and orientation. The work in [15] describes that the place cells firing are driven by the orientation and speed perceived by head direction cells and proposes a goal-directed spatial navigation model using forward trajectory

planning. As known from these studies, the progress of brain science is affecting the developmental direction of navigation technology and it is feasible to solve the problems about navigation by making use of the discoveries in the field of brain science. However, for the issue of simulating the biological place cell to realize positioning, the above-mentioned studies on building the model of spatial representation are all based on the single scale of place cells map, which is not in line with the biological basis[16] and does not have a true reflection on how the location information drives the place cell firing. The latest research data shows that the location in space is determined by the results of the comprehensive characterization of place cells with different scales in the hippocampus[16],[17]. Thus, The authors consider how to construct a spatial cognitive model of multi-scale place cells to achieve goal-oriented navigation, which is not only more consistent with the biological basis, but also can better reflect the biological capacity of dealing effectively with the perception of space information.

In this paper, a goal-oriented navigation model based on a multi-scale place cells map is proposed. A multi-scale place cell map is constructed to characterize the location of the robot. The reinforcement learning mechanism is introduced to improve the understanding of the robot for the space environment and complete the goal-oriented navigation task successfully. The rest of the article is organized as follows. Section 2 introduces the Q-learning algorithm and the model of the place cell firing rate, and then focuses on the proposed multi-scale spatial representation model. Section 3 is the simulation and discussion on the proposed method. Section 4 summarizes the whole paper.

## 2. Materials and Methods

### 2.1 Q-learning Algorithm

Q-learning is one of the reinforcement learning algorithms[18] which uses numerical iterative calculations to approximate the optimal value. The algorithm first initializes a table of Q value, and chooses an action according to the Q value in current state, then updates the corresponding Q value according to the feedback of the reward. If the action receives a positive reward, the corresponding Q value will increase and the robot will be inclined to exact the operation in future. While if the action receives a negative reward, the corresponding Q value will decrease and the robot will not be inclined to exact the operation. Through several episodes, the action will tend to the optimal policy.  $Q(s,a)$  represents the value function of performing an action  $a$  in state  $s$ . In the Q-learning process, the optimal value function can be approximated by iterative calculations. The update rules are as follows:

$$Q(s,a) = R + \gamma \max_{a'} Q(s',a') \quad (1)$$

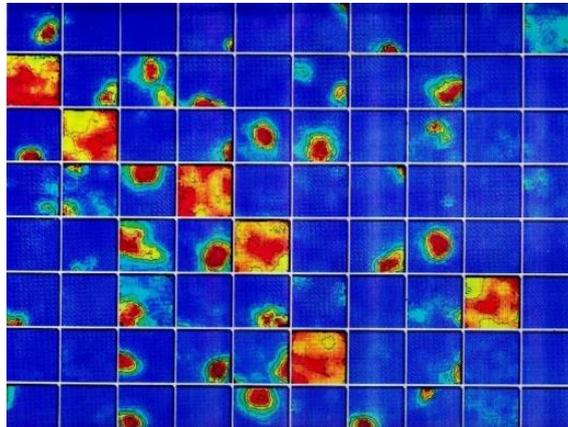
$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha \left( R + \gamma \max_{a'} Q(s',a') \right) \quad (2)$$

Where  $R$  is the reward function, and  $\gamma$  is the discount factor,  $\gamma \in [0,1)$ .  $\alpha$  is the learning rate which controls the speed of learning. The greater the value of  $\alpha$ , the faster the convergence. But too large  $\alpha$  may cause premature convergence. So it is important for learning speed to choose a proper  $\alpha$ .

### 2.2 The Model of Place Cells Firing

O'Keefe finds that there are some special neural cells in rat's hippocampus. Their firing activities are closely related to the spatial location of the rat[19]. When the rat is in a relatively narrow range, if some cell shows firing with a maximum frequency while in

other places with little or no firing, then the cell is called place cell by O'Keefe, and the corresponding range of environment of the firing cell is called place field. This particular firing activity of the cell helps animal determine its position in the current environment. A single place cell corresponds with a location in the environment. A particular environment can be characterized by the electrical activities of a number of place cells. Figure 1 shows a response diagram of place cells firing in CA3[20]. It can be seen that there are some corresponding place cells firing when the rat moves in the environment and the intermediate region of place cell has a maximum response while decreasing outwards in turn.

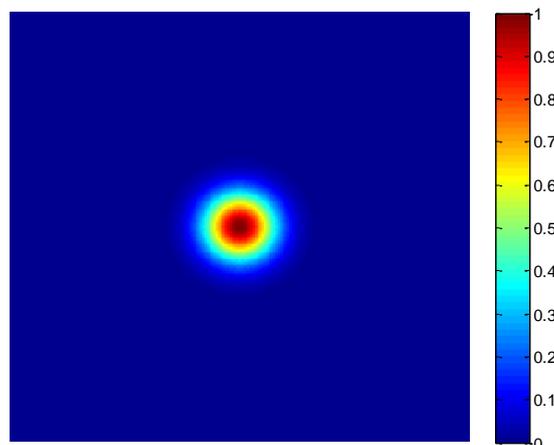


**Figure 1. A Response Diagram of Place Cells Firing[20]**

According to the shape of the response diagram when place cells are activated in space, O'Keefe proposes to use Gaussian function to make a simulation. Let  $(x, y)$  is the spatial position of the vehicle, the firing rate of the place cell  $i$  can be defined by:

$$f r_i(x, y) = \exp \left\{ -\frac{(x-x_i)^2 + (y-y_i)^2}{\sigma_0^2} \right\} \quad (3)$$

Where  $(x_i, y_i)$  is the reference position of  $P C_i$ , and  $\sigma_0$  is the adjustable factor of place field. Figure 2 shows the distribution of a place cell in firing state while  $\sigma_0^2 = 50$ . From the figure 2, it can be seen that the simulated place cell has a maximum response in the intermediate red area where the firing rate in the center is equal to 1, and the firing rate in the external blue color area is close to 0. Therefore, Eq.(3) is able to simulate the characteristics of the place cell in hippocampus to some extent.



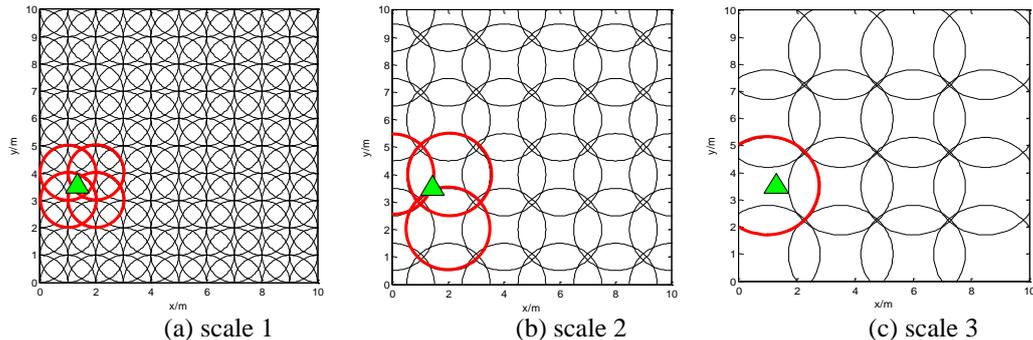
**Figure 2. A Schematic Diagram of the Simulated Place Cell**

### 2.3 The Multi-scale Spatial Representation for Goal-oriented Navigation

Geva-Sagiv pointed out in the latest study that the hippocampus of the bats and mice can build multi-scale spatial cognitive map according to the size and complexity of the environment and complete autonomous navigation[16]. In the moving process, the place cells in firing state exhibit different sizes and the scales of place field decrease gradually along the dorsal to the ventral[21]. As these place cells under different scales discharge simultaneously, the location of the robot can be characterized by the place cells in firing state at different scales. The reinforcement learning mechanism is introduced in order to finish goal-oriented navigation. The main steps are as follows:

**Step1 (Building a multi-scale place cells map):** The firing of place cells is driven by external sensory and self-motion information. This paper focuses on how to make use of the firing effect of place cells to give the robot a reasonable action and complete the goal-oriented navigation. Therefore, The authors consider that this stimulation process has finished and use the simulated location information directly drives place cells fire. The Gaussian model introduced in Section 2.2 is adopted to calculate the firing rate of place cells in the subsequent analysis process.

Suppose the moving area of the robot is covered by a  $a' \times a$  square and the place cells are built uniformly at a certain interval. Thus, any position of the robot in the environment is related to the corresponding place cell. As there may be several place cells in firing state simultaneously, the place cells map under different scales can be formed by modulating the size of place field. Figure 3 shows a schematic diagram of place cells map in three different layers in which the black circles indicate the constructed place cells, the green triangle indicates the robot and the red circles indicate the place cells in firing state. As can be seen from the figure 3, the robot in current location has corresponding firing place cells in different layers, where the small scale has four place cells in firing state, the medium scale has three place cells in firing state and the large scale has one place cell in firing state. Thus, the position of the robot in the environment can be synthetically characterized by the firing place cells in different layers.



**Figure 3. A Multi-scale Place Cells Schematic**

**Step2 (Rewards function design):** The goal-oriented navigation task requires the robot not to encounter the obstacles from the initial state to the target in the moving process and to plan a short path. So the design of the rewards function should timely and accurately reflect the effect of implement different actions in each state. Thus, the rewards function is designed from two viewpoints of reaching the target state and colliding with the obstacles.  $d_{goal}(t)$  represents the distance from the robot to the target,  $d_0$  is the maximum target detection range,  $d_{obs}(t)$  represents the distance between the robot and the obstacle,  $d_{min}$  is the minimum dangerous distance,  $d_{max}$  is the maximum safe distance. So the rewards function  $R$  can be expressed as:

$$R = \begin{cases} 10 & d_{goal}(t) < d_0 \\ -10 & d_{obs}(t) < d_{min} \\ 2 & d_{min} < d_{obs}(t) < d_{max} \quad \text{and} \quad d_{obs}(t-1) < d_{obs}(t) \\ -2 & d_{min} < d_{obs}(t) < d_{max} \quad \text{and} \quad d_{obs}(t-1) \geq d_{obs}(t) \\ 0 & \text{others} \end{cases} \quad (4)$$

When the robot runs into the target detection range, a large positive reward value is given. For collision avoidance behavior, when the distance between the robot and the obstacle in next moment is less than the minimum dangerous distance, a negative positive reward value is given, and when the distance between the robot and the obstacle in next moment is between the minimum dangerous distance and the maximum safe distance, a small reward value is given according to the robot close to or away from the obstacle in next moment.

**Step3 (Action selection policy):** In the phase of action selection, there is a common problem on the balance between exploration and exploitation. In the beginning, the agent has no prior information about the environment and the task, so it is inclined to select an action randomly. After several trials, the agent accumulates some experience and tends to select an optimal action. In this article, the method of simulated annealing is adopted to execute action selection. The probability of perform an action  $a$  under the state  $s$  is defined by Eq.(5):

$$P(a | s) = \frac{\exp[Q(s, a)/T]}{\sum_{b \in A} \exp[Q(s, b)/T]} \quad (5)$$

$$T_{k+1} = \lambda^k T_0 \quad (6)$$

Where  $T$  is the temperature parameter. The geometric cooling strategies is used according to Eq.(6).  $T_0$  is the initial value of temperature parameter.  $\lambda$  is the cooling parameter ( $0 < \lambda < 1$ ).

At the beginning of learning, temperature control parameter  $T$  is large, and the probability of all actions is almost equal, in which the robot is inclined to explore. With the learning episodes increasing, the value of  $T$  gradually decreases, while the probability of random action also reduces and the better action will be favored. In the latter learning, the value of  $T$  is closely reduced to 0 when this phase is equivalent to adopting greedy strategy and the optimal action is chosen.

**Step4 (Q-learning process):** For the constructed place cells map in each layer, first of all, the firing rate of place cells should be calculated according to Eq.(3). Since there may be several different place cells in firing state at the same time, a comparison is drawn in order to find the place cell with maximum firing intensity. So the current location can be represented by this place cell. Then we should consider how to choose a reasonable action to move to the next location. In current state, the robot has several actions to choose. The Q values of different actions are compared and the action with maximum Q value is identified as the output. So an action is selected in current state. The whole process can be seen as Eq.(7).

$$a = \arg \max_a \left\{ Q(s_i, a) \cdot \max_i [fr_i(x, y)] \right\} \quad (7)$$

According to Eq.(7), an action will be given by the place cells map in each layer. The integration of different scales of map helps to make a judgment and decision on action selection. Then the method of simulated annealing is adopted so as to keep the balance

between exploration and exploitation. Once an action  $a$  is executed, the Q values in each layer should be updated according to the action  $a$ . After continuous iterative learning, the robot is able to integrate the behavioral information effectively and give a reasonable action to plan out an optimal route to the goal. Compared with the classical Q-learning algorithm, the method in this paper adds the firing activities of place cells to the algorithm, which creates a bridge between the firing rate of place cells and the actions and is able to help the robot to plan an optimal route from the starting point to the goal. If there is only one place cell in firing state in current location, the activities of different place cells are unessential to be calculated. Thus, the proposed method degenerates to the classical Q-learning algorithm.

Figure 4 shows a flow chart of the method in one episode. First, build a multi-scale place cells map and calculate the firing rate of place cells in different layers. Second, confirm the current state according to the firing intensity. Third, select an appropriate action and move the robot to the next location. Fourth, update the Q values and give a judgment whether the robot reaches the detection range of the goal. If it is, an exploration ends, if not, the robot continues to perform the same steps in new state. After several cycles of exploration, the robot can accumulate the experience knowledge and learn a suitable navigation policy to help itself move to the goal with an optimal route.

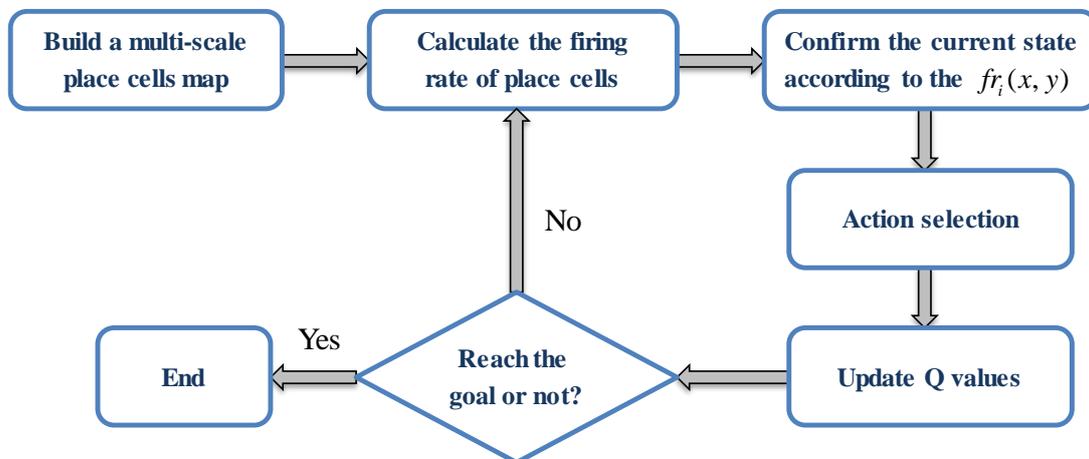


Figure 4. A Flow Chart of the Method

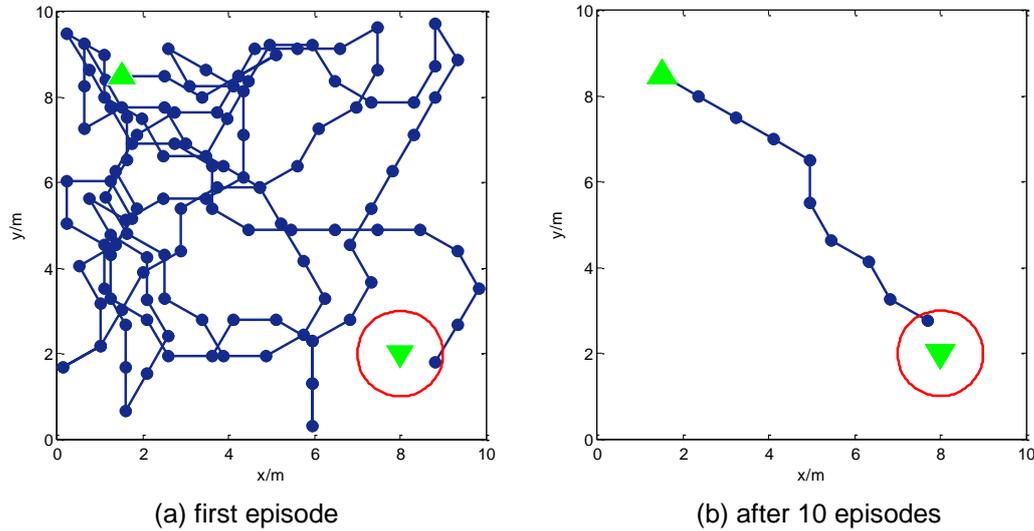
### 3. Results and Discussion

In order to demonstrate the feasibility and advantages of the method for completing the task of goal-directed navigation, a detailed analysis is made in simulation. All simulations are based on the platform of Matlab R2010a. First of all, the parameters involved in the simulation are set and introduced as follows.

(1) Space Environment: it assumes that the robot moves in a square area with the size of 10m×10m. (2) The scale of place cells: the multi-scale place cells map is composed of three different scales in simulation, in which the radius of place fields is respectively set to 0.8m, 1.4m, 2m. (3) The starting point and the destination point are set to (1.5,8.5) and (8,2). The target detection range is a circular region of which radius is 1m, (4) Action settings: there are five actions to choose for the agent, respectively, rotate  $\pi/6$  to the left, rotate  $\pi/3$  to the left, no rotation, rotate  $\pi/6$  to the right, rotate  $\pi/3$  to the right, and for each action the agent goes forward 1m along the current direction after rotation. (5) The learning rate and the discount factor are respectively set to  $\alpha=0.7$ ,  $\gamma=0.9$ . (6) The minimum dangerous distance  $d_{\min}$  is 0.1m and the maximum safe distance  $d_{\max}$  is

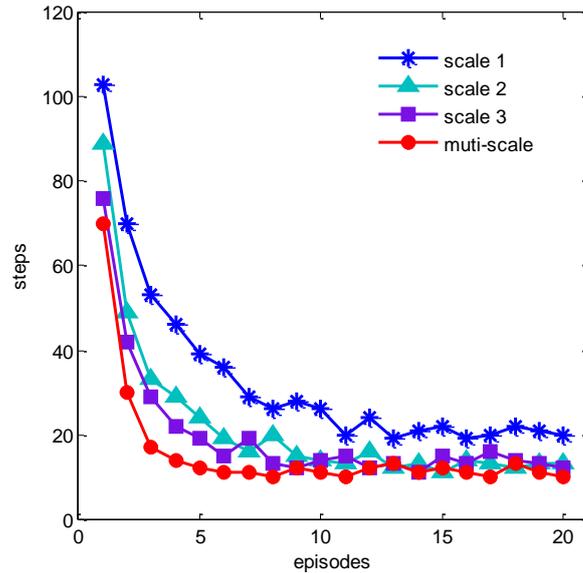
1m. When the agent runs into the range, an episode is considered to finish. After training and learning, the algorithm can plan an optimal route from the starting point to the goal.

Figure 5 shows two planned paths from the starting point to the goal in the first episode and after 10 episodes in which the green triangle represents the starting point, the green del triangle represents the goal and the red circle represents the target detection range. As can be seen from the figure 5, in the first episode, the robot often moves randomly because of lack of priori information about the environment and tasks. After 10 episodes, the cumulative experience has a good guide for the robot and an optimal route to the goal is planned successfully. The simulation results confirm that it is feasible for the proposed method to complete the goal-oriented navigation.



**Figure 5. Path Planning before Learning and after Learning**

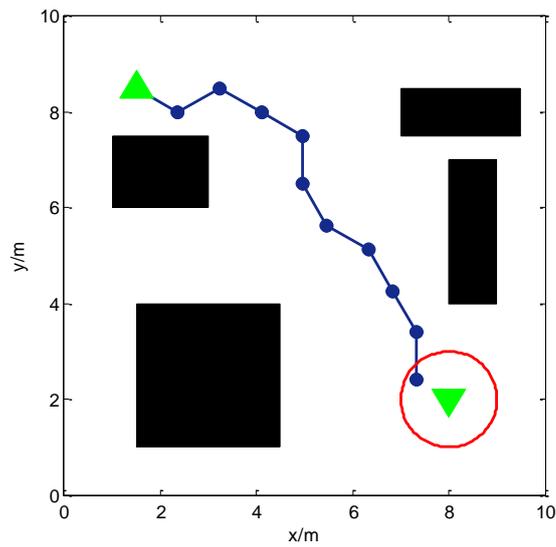
To illustrate the advantages of the method for goal-oriented navigation, a contrastive analysis is made in simulation between the single scale spatial representation with three different scales and the multi-scale spatial representation. The scales are respectively set as  $scale\ 1 = 2m$ ,  $scale\ 2 = 1.4m$ ,  $scale\ 3 = 0.8m$ , and other parameters are same as the above. Figure 6 shows the average number of required steps from the starting point to the target with the variation of episodes. The experiment under each episode is carried out for 100 times repeatedly and then the average steps are calculated.



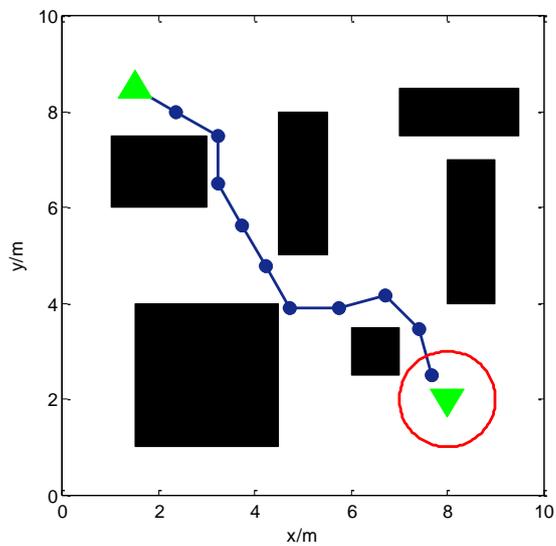
**Figure 6. The Average Number of Required Steps with the Variation of Episodes**

As can be seen from the figure 6, with the number of episodes increasing, the required steps to the target rapidly decrease. And when the episodes increase to a certain value, the number of steps is not changed, which represents the simulated robot has found the shortest path to the destination. For the large scale place cells map, the agent needs more steps and can not plan a shortest path to the goal. For the medium scale and the small scale, the agent is able to find a optimized through learning. For the multi-scale place cells map, the agent needs fewer number of steps under the condition of same episodes and the learning speed is faster. Thus, the proposed method of the multi-scale spatial representation has a stronger ability of learning to complete the goal-oriented navigation.

In order to test the feasibility of the proposed method in the environment with obstacles, some different obstacles are randomly set in the simulated area and a test is made to observe whether the agent can find a collision-free route to the goal . As can be seen from the figure 7, the agent is able to plan an optimal path without colliding with the obstacles in the environment. Furthermore, the simulated environment is changed and another two obstacles are put in the space as the figure 8 shows. The agent can give a reasonable action to avoid a collision when confronted with new obstacles and succeed to reach the goal. Therefore, this method is feasible for goal-oriented navigation in the presence of obstacles in the environment. And when the obstacles change, the method is still effective.



**Figure 7. The Goal-oriented Navigation in the Environment with Obstacles**



**Figure 8. The Goal-oriented Navigation in the Environment with Changed Obstacles**

#### 4. Conclusion

Enlightened by the multi-scale nature of spatial representation in biology, a goal-oriented navigation model based on a multi-scale place cells map is proposed in this article. Reinforcement learning algorithm is used to build a relation between the firing rate of place cells and the actions. Through training and learning, the agent can gain experience knowledge and plan the shortest path from the starting point to the goal. Simulation results show that the proposed multi-scale spatial cognition model contributes to faster learning and is better able to complete the task of goal oriented navigation in the environment with obstacles. In future work, the authors will focus on

improving the details of the navigational model and testing the performance of the method in real platform.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant No.61273048.

## References

- [1] E.C. Tolman. Cognitive maps in rats and men. *Psychological Review*. 1948, 55,189-208.
- [2] J. O'Keefe and L. Nadel. *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press. 1978.
- [3] G. Brett, N.B. William and S. Jeffery. The Head-Direction Signal Is Critical for Navigation Requiring a Cognitive Map but Not for Learning a Spatial Habit. *Current Biology*. 2013, 23, 1536-1540.
- [4] R. Florian, P.B. Mark, G.C. William and E.H. Michael. Head direction is coded more strongly than movement direction in a population of entorhinal neurons. *Brain Research*. 2015, 1621, 355-367.
- [5] B. Daniel, B. Caswell, M. Daniel and N. Burgess. Using Grid Cells for Navigation. *Neuron*, 2015, 87, 507-520.
- [6] B. Yoram. Spatial coding and attractor dynamics of grid cells in the entorhinal cortex. *Current Opinion in Neurobiology*. 2014, 25, 169-175.
- [7] T. Hartley, C. Lever, N. Burgess and J. O'Keefe. Space in the brain: how the hippocampal formation supports spatial cognition. *Philosophical Transactions of the Royal Society B Biological Sciences*. 2013, 369(1635), 20120510.
- [8] K. Emilio, E.C. James, M.B. Moser and E.I. Moser. Speed cells in the medial entorhinal cortex. *Nature*. 2015, 523(7561), 419-424.
- [9] W. Thomas. Spatial Navigation. *International Encyclopedia of the Social & Behavioral Sciences*. 2015, 23, 161-171.
- [10] C. Yan, R.B. Wang and J.Y. Qu. Locating and navigation mechanism based on place-cell and grid-cell models. *Cognitive Neurodynamics*. 2016,1-8.
- [11] H.F. Ólafsdóttir, C. Barry and A.B. Saleem. Hippocampal place cells construct reward related sequences through unexplored space. *Elife Sciences*, 2015, 4, 1-17.
- [12] J. Metz, T. Szilágyi and M. Perian. Computer modeling of hippocampal CA1 pyramidal cells - a tool for in silico experiments. *Acta Medica Marisiensis*. 2015, 61(1), 15-24.
- [13] B. Alejandra, T. Gonzalo, L. Martin and W. Alfredo. Learning Spatial Localization: From Rat Studies to Computational Models of the Hippocampus. *Spatial Cognition & Computation: An Interdisciplinary Journal*. 2015, 15, 27-59.
- [14] J.S. Hugo and B. Caswell. Neural systems supporting navigation. *Current Opinion in Behavioral Sciences*. 2015(1), 47-55.
- [15] U.M. Erdem and E.H. Michael. A biologically inspired hierarchical goal directed navigation model. *Journal of Physiology-Paris*. 2014, 108, 28-37.
- [16] G.S. Maya, L. Lroia, Y. Yossi and U. Nachum. Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation. *Nature Reviews Neuroscience*. 2015, 16(2), 94-108.
- [17] M. Llofriu, G. Tejera, M. Contreras, T. Pelc, J.M. Fellous and A. Weitzenfeld. Goal-oriented robot navigation learning using a multi-scale space representation. *Neural Networks*. 2015, 72, 62-74.
- [18] A. Nihal. Reinforcement learning-based mobile robot navigation. *Turkish Journal of Electrical Engineering & Computer Sciences*. 2016, 24, 1747-1767.
- [19] K. Jeffery. Spatial Cognition: Entorhinal Cortex and the Hippocampal Place-Cell Map. *Current Biology*. 2015, 25(24), 1181-1183.
- [20] J. O'Keefe and N. Burgess. Geometric determinants of the place fields of hippocampal neurons. *Nature*, 1996, 381(6581), 425-428.
- [21] K.B. Kjelstrup, T. Solstad, V.H. Brun, T. Hafting, S. Leutgeb, M.P. Witter, E.I. Moser and M.B. Moser. Finite scale of spatial representation in the hippocampus. *Science*. 2008, 321(5885), 140-143.

## Authors



**Jia Du**, She received her M.Sc. degree in 2008 from Xi'an Communications Institute. Now, she is a Ph.D. candidate at Air Force Engineering University. Her main research directions include intelligent and autonomous navigation, biological trajectory planning.



**Dewei Wu**, He is a professor and Ph.D. supervisor at Air Force Engineering University now. His main research directions include navigation positioning theory, technology and application, intelligent and autonomous navigation.



**Weilong Li**, He received his M.Sc. degree in 2012 from Air Force Engineering University. Now, he is a Ph.D. candidate there. His main research directions include intelligent and autonomous navigation, biological computer vision.



**Yang Zhou**, He received his M.Sc. degree in 2014 from Air Force Engineering University. Now, he is a Ph.D. candidate there. His main research direction is intelligent and autonomous navigation.

